UWM-HBUT at TREC 2014 Microblog Track: Using Query Expansion (QE) and Event Identification Algorithm (EIA) to improve microblog retrieval effectiveness

Sukjin You
University of Wisconsin-Milwaukee
P.O. Box 413,
Milwaukee, WI 53211
yous@uwm.edu

Wei Huang[*]
Hubei University of Technology Wuhan,
China 430068
tonny_hw@hotmail.com

Xiangming Mu
University of Wisconsin-Milwaukee
P.O. Box 413,
Milwaukee, WI 53211
mux@uwm.edu

ABSTRACT

This paper reports our contributions and results to TREC 2014 Microblog Track. Different from traditional web pages or database documents, microblogs have their own unique features. Considering sensitivity to time, we introduce a new factor to help to improve tweet retrieval effectiveness. The ranking score of a retrieved tweet is adjusted by considering how close the tweet time stamp is to the event using Event Identification Algorithm (EIA). In addition, we also evaluate the Query Expansion (QE) approach using Google as an external data corpus. There are 55 search topics and the data set contains a total of 243 million tweets provided by the TREC 2014 Microblog Track. Our initial results indicated that QE helped to improve the performance. We also discussed why the EIA approach failed to enhance the retrieval performance.

Keywords: performance evaluation, twitter, microblog retrieval, query expansion, event identification algorithm

1. INTRODUCTION

Twitter is one of the most popular microblog platforms. Microblogs help people to share messages and interact with their social relations in a convenient way and it becomes a valuable information source in addition to the traditional web and database documents. Today there are 230 million active users on Twitter and they post an average of 500 million Tweets a day. Tweets being posted are usually short (less than 140 characters) and time sensitive. Tweets include more social chats and social events, while traditional web pages contain more basic facts and navigational contents. It is interesting to explore new approaches for effective microblog information retrieval. A summary of issues and recent research progresses regarding this topic is provided by Efron (2011).

The 2014 Microblog Track has two tasks: temporally-anchored ad hoc retrieval and tweet timeline generation (TTG). In this study we used data from the TREC 2014 Microblog Track and completed the first task. The task is to retrieve relevant tweet documents for each provided query. We tested the effectiveness of a new weighted Query Expansion approach. Considering the time sensitivity feature of microblogs, we also propose to adjust weighting factors for the tweet rank based on our model using an event detection algorithm.

2. RELATED WORK

Recent Query Expansion (QE) methods for microblog search utilize temporal properties derived from the real-time characteristic that many messages are posted by users when an interesting event has occurred. Massoudi (2011) presented that QE on microblog data can be done in a dynamic fashion (taking time into account) and should include specific terms like usernames, hashtags, and links. Louvan (2011) found the QE utilizing external search results combined with re-tweet value in the customized scoring function was the most effective. Metzler (2012) proposed an approach that is unsupervised in the sense that it makes use of a pseudo-relevance feedback-like mechanism when extracting expansion terms. Miyanishi (2013) indicated that by retrieving useful information from among a huge quantity of authors' messages, QE is used to enrich a user query. This

[*] Wei Huang is the corresponding author and also a visiting scholar of University of Wisconsin at Milwaukee.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **NOV 2014**	2. REPORT TYPE	3. DATES COVERED **00-00-2014 to 00-00-2014**

4. TITLE AND SUBTITLE
UWM-HBUT at TREC 2014 Microblog Track: Using Query Expansion (QE) and Event Identification Algorithm (EIA) to Improve Microblog Retrieval Effectiveness

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Wisconsin-Milwaukee,PO Box 413,Milwaukee,WI,53211

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
presented in the proceedings of the Twenty-Third Text REtrieval Conference (TREC 2014) held in Gaithersburg, Maryland, November 19-21, 2014. The conference was co-sponsored by the National Institute of Standards and Technology (NIST) and the Defense Advanced Research Projects Agency (DARPA).

14. ABSTRACT
This paper reports our contributions and results to TREC 2014 Microblog Track. Different from traditional web pages or database documents, microblogs have their own unique features. Considering sensitivity to time, we introduce a new factor to help to improve tweet retrieval effectiveness. The ranking score of a retrieved tweet is adjusted by considering how close the tweet time stamp is to the event using Event Identification Algorithm (EIA). In addition, we also evaluate the Query Expansion (QE) approach using Google as an external data corpus. There are 55 search topics and the data set contains a total of 243 million tweets provided by the TREC 2014 Microblog Track. Our initial results indicated that QE helped to improve the performance. We also discussed why the EIA approach failed to enhance the retrieval performance.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**5**	

approach leverages temporal properties for QE and combines them according to the temporal variation of a given topic. Experimental results show that this QE method using automatically combined temporal properties is effective at improving retrieval performance.

Features of microblogs are used to detect the relevance to the query. Zhao (2011) focused on six factors of microblogs: the number of total tweets a user posts, the number of total followers a user has, the length of a tweet a user posts, the number of URLs a tweet contains, the number of retweets in a tweet, and the number of mentions in a tweet. They measured a user's social influence, and each of them is highly relevant to the social network properties of the microblog authors and the properties of the microblog itself. Damak (2013) used 14 selected features to learn models to determine their effectiveness in a microblog search task. These features are evaluated in a context of microblog information retrieval. They found that tweet popularity, length, exact term matching, URL presence, URL popularity in the corpus, URL frequency in the microblog and the recency of the microblog are the features that are best connecting with the judgment of relevance.

On the condition of multi-feature of microblog, faceted search was connected to microblog search. Facets are defined as "a set of meaningful labels organized in such a way as to reflect the concepts relevant to a domain" by Hearst (2006, p. 60). Abel (2011) proposed strategies for inferring facets and facet values on Twitter by enriching the semantics of individual Twitter messages (tweets) and by presenting different methods, including personalized and context-adaptive methods. Vosecky (2013) defined the Multi-Faceted Topic Model (MfTM) which is proposed to jointly model latent semantics among terms and entities and capture the temporal characteristics of each topic. According to the author, this approach could capture the dynamic and entity-oriented topics in microblogs.

Microblog queries usually can be classified as: celebrity, social event and common queries (Teevan, Ramage, & Morris, 2011). The topic detection and clustering of microblogs are valuable methods used in microblog search. O'Connor (2010) completed the topic extraction system based on syntactic filtering, language modeling, near-duplicate detection, and set cover heuristics. Long (2011) proposed a unified workflow of event detection, tracking and summarization on microblog data. A bipartite graph is constructed to capture the relationship between two events occurring at adjacent times. The matched event pair is grouped into an event chain. Furthermore, inspired by diversity theory in Web search, the authors summarized event chains by considering the content coverage and evolution over time. Huang (2012) adopted the Single-pass Clustering technique by using the Latent Dirichlet Allocation (LDA) model in place of the traditional VSM model, to extract the hidden microblog topic information. Chen (2013) designed a novel scheme to crawl through the relevant messages related to the designated organization by monitoring multi-aspects of microblog content, including users, the evolving keywords and their temporal sequence, then developed an incremental clustering framework to detect new topics, employing a range of contents and temporal features to help in promptly detecting hot emerging topics.

Ranking and re-ranking of tweets, as search results for a query is challenging, among other things because of the sheer amount of microblogs that are being generated in real time, as well as the short length of each individual microblog. Amati (2011) proposed Time Re-Ranking Component, which assumed that time and relevance are two dependent variables. Time and relevance yield two independent rankings, which need to be merged. This approach based on a comparison of score and timestamps of the retrieved messages. The main idea is to choose a particular score value as a filtering threshold. Miyanishi (2011) developed a two-step approach: re-ranking and filtering. The filtering step identifies and removes retweets and non-English tweets, which was found to be crucial for this task. After filtering, tweets are re-ranked based on the Learning-to-Rank (L2R) model learned using five types of features. The re-ranking step further improved the search performance, achieving the best overall result. Roegiest (2011) used feature engineering and a variety of ranking methods, and combined them using Reciprocal-Rank-Fusion (RRF) as a method of meta-ranking the results.

3. METHODOLOGY

TREC microblog track 2014 dataset provided by NIST was collected via the Twitter streaming API over a two-month period: 1 February, 2013 - 31 March, 2013 (inclusive). The collection consists of approximately 243 million tweets (statuses).
In this study we conducted a total of four runs:

- **Run UWMHBUT1**

A Base run using TREC API (RunQueriesThrift.java). Some result records were duplicated. We removed duplicates, and added 1001th records for these topics (MB178, MB179, MB189, MB194, and MB197). This TREC API was developed based on the Lucene project using language model.

- **Run UWMHBUT2**

A Query Expansion approach based on the term frequency from the first top 10 Google results with weight for each term. The title and abstract of the top 10 Google result items were used to calculate term frequency. After the stop words were removed,

nine highest term frequency terms were added to the original query: $T_{i,}$ (i=0…C-1, C=10, T_0 = original query). Weight w_j was given to each term T_j:
$w_j = (C\text{-}j) / \sum_{=0}^{-1}(i+1)$, j=0…C-1

Where C=10.

This approach is based on Kwok et al.'s (2005) idea by introducing web assistance for improved performance.

- **Run UWMHBUT3**

Same as previous QWR (UWMHBUT2), but with equal weight for each expanded term.

- **Run UWMHBUT4**

Event Identification Algorithm: The distribution of twitter records in a time period is assumed to be a Gaussian distribution. We assume that the mean value of that distribution also represents the maximum tweets posted about the topic. The time of search results is used to adjust their rankings. The top 30 search results were analyzed and the mean of the time posted was calculated. More weightage was given to tweets posted closer to the mean. For an example, when topic "Mad men season 6" is searched we assume that maximum number of tweets were posted regarding this topic when the season 6 was launched. In this case more weight was given to tweets posted during that period.

Specifically,
$R_n = R_o (\alpha + (1\text{-}\alpha) E)$,

where, R_n is the new ranking score, R_o is the original ranking score calculated from TREC Microblog APIs, and E is the event effect. Here α is adjusting parameter (we choose 0.8 for this run). The event weighing factor E is calculated by the following formula:

$E = f(t, \mu, \sigma) * R_o$

where, $f(t, \mu, \sigma)$ is the Gaussian distribution, t is the time gap between the querytime and the time where the twitter record was posted, μ is the event center (or the "hot" point), and σ is the standard deviation. We calculated the σ based on the top 1500 search results. To smooth it, we use $\sigma = 3 * \sigma_1$ were σ_1 is the standard deviation of the top 1500 search results in terms of days.

4. RESULTS

Table 1: Results from four runs (UWMHBUT1,2,3,4) using MAP and P@30

RUNS	MAP	P@30
UWMHBUT1	0.309	0.515
UWMHBUT2	$0.343^{\dagger}$	0.512
UWMHBUT3	0.336	0.496
UWMHBUT4	0.308	0.512

† means difference with baseline (UWMHBUT1) is statistically significant

Table 1 presents the MAP and P@30 scores for our four runs: UWMHBUT1, UWMHBUT2, UWMHBUT3 and UWMHBUT4. Using a T-test, we found the MAP score on UWMHBUT2 is significantly better than the baseline (UWMHBUT1) MAP score (df=54, p=0.025).

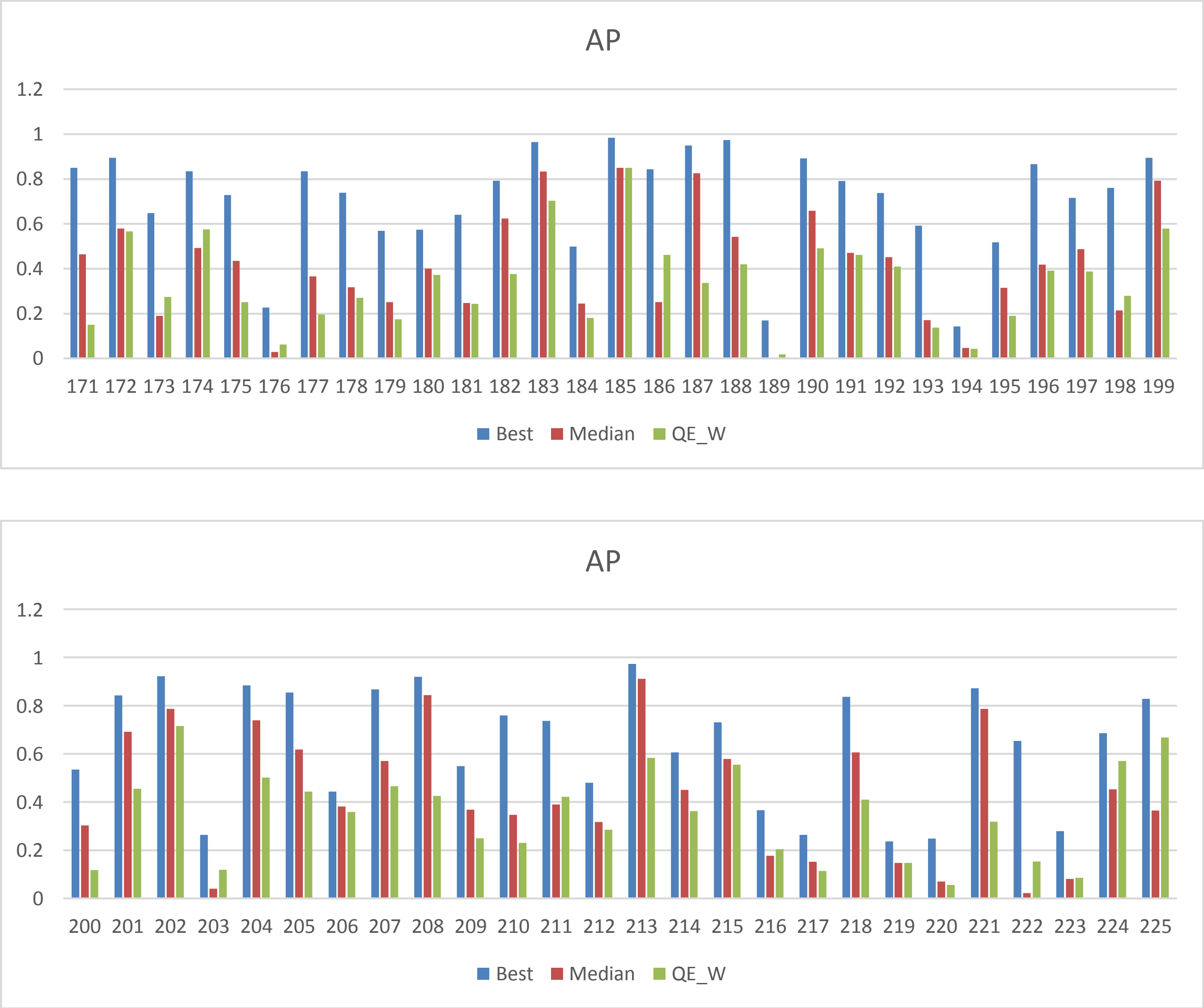

Figure 1: Average Precision comparison: BEST, MEDIAN, and QUERY EXPANSION (WEIGHTED)

5. CONCLUSIONS

Weighted QE based on external data corpus (here we used Google collection) approach (UWMHBUT2) provided improved MAP score as compared to the baseline.

Event Identification Algorithm (EIA) was not found to be helpful in terms of both MAP and P@30 scores. One possible explanation is the data collection is from a short period of time (two months).

6. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 71303075) and China Scholarship Council.

7. REFERENCES

Abel, F., Celik, I., Houben, G. J., & Siehndel, P. (2011). Leveraging the semantics of tweets for adaptive faceted search on twitter. In The Semantic Web–ISWC 2011 (pp. 1-17). Springer Berlin Heidelberg.

Amati, G., Amodeo, G., Bianchi, M., Marcone, G., Bordoni, F. U., Gaibisso, C., ... & Flammini, M. (2011). FUB, IASI-CNR, UNIVAQ at TREC 2011 Microblog Track. In TREC.

Chen, Y., Amiri, H., Li, Z., & Chua, T. S. (2013, July). Emerging topic detection for organizations from microblogs. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (pp. 43-52). ACM.

Damak, F., Pinel-Sauvagnat, K., Boughanem, M., & Cabanac, G. (2013, March). Effectiveness of State-of-the-art Features for Microblog Search. InProceedings of the 28th Annual ACM Symposium on Applied Computing (pp. 914-919). ACM.

Efron, M. (2011). Information search and retrieval in microblogs. Journal of the American Society for Information Science and Technology, 62(6), 996-1008.

http://trec.nist.gov/tracks.html.

https://business.twitter.com/whos-twitter.

Hearst, M. A. (2006). Clustering versus faceted categories for information exploration. *Communications of the ACM*, *49*(4), 59-61.

Huang, B., Yang, Y., Mahmood, A., & Wang, H. (2012, January). Microblog topic detection based on LDA model and single-pass clustering. In Rough Sets and Current Trends in Computing (pp. 166-171). Springer Berlin Heidelberg.

Kwok, K. L., Grunfeld, L., & Deng, P. (2005). Improving weak ad-hoc retrieval by web assistance and data fusion. In *Information Retrieval Technology* (pp. 17-30). Springer Berlin Heidelberg.

Li, R., Wei, B., Lu, K., & Wang, B. (2011). Author Model and Negative Feedback Methods on TREC 2011 Microblog Track. In TREC.

Long, R., Wang, H., Chen, Y., Jin, O., & Yu, Y. (2011). Towards effective event detection, tracking and summarization on microblog data. In Web-Age Information Management (pp. 652-663). Springer Berlin Heidelberg.

Louvan, S., Ibrahim, M., Adriani, M., Vania, C., Distiawan, B., & Wanagiri, M. Z. (2011). University of Indonesia at TREC 2011 microblog track. In Text REtrieval Conference Proceedings. NIST.

Louvan, S., Ibrahim, M., Adriani, M., Vania, C., Distiawan, B., & Wanagiri, M. Z. (2011). University of Indonesia at TREC 2011 microblog track. In Text REtrieval Conference Proceedings. NIST.

Massoudi, K., Tsagkias, M., de Rijke, M., & Weerkamp, W. (2011). Incorporating query expansion and quality indicators in searching microblog posts. In Advances in Information Retrieval (pp. 362-367). Springer Berlin Heidelberg.

Metzler, D., Cai, C., & Hovy, E. (2012, June). Structured event retrieval over microblog archives. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 646-655). Association for Computational Linguistics.

Miyanishi, T., Seki, K., & Uehara, K. (2012). TREC 2012 microblog track experiments at Kobe University. KOBE UNIV (JAPAN).

Miyanishi, T., Seki, K., & Uehara, K. (2013). Combining recency and topic-dependent temporal variation for microblog search. In Advances in Information Retrieval (pp. 331-343). Springer Berlin Heidelberg.

Nagmoti, R., Teredesai, A., & De Cock, M. (2010, August). Ranking approaches for microblog search. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on (Vol. 1, pp. 153-157). IEEE.

O'Connor, B., Krieger, M., & Ahn, D. (2010, May). TweetMotif: Exploratory Search and Topic Summarization for Twitter. In ICWSM.

Roegiest, A., & Cormack, G. V. (2011). University of Waterloo at TREC 2011 Microblog Track. In TREC.

Teevan, J., Ramage, D., & Morris, M. R. (2011, February). # TwitterSearch: a comparison of microblog search and web search. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 35-44). ACM.

Vosecky, J., Jiang, D., Leung, K. W. T., & Ng, W. (2013, October). Dynamic multi-faceted topic discovery in twitter. In Proceedings of the 22nd ACM international conference on Conference on information & knowledge management (pp. 879-884). ACM.

Zhai, C., & Lafferty, J. (2001, September). A study of smoothing methods for language models applied to ad hoc information retrieval. In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 334-342). ACM.

Zhao, L., Zeng, Y., & Zhong, N. (2011). A weighted multi-factor algorithm for microblog search. In Active Media Technology (pp. 153-161). Springer Berlin Heidelberg.